In examining what behaviors and characteristics mostly likely present a diagnosis of Autism, I started by looking at summary statistics for characteristics at a high level by general category such as age (in months), gender, ethnicity, etc. I created a field for age range, putting age into five buckets.

At this point, I did not look at the individual behavior observations recorded in fields A1 through A10. The attributes of these fields are as follows.

A1-A10: Items within Q-Chat-10 in which questions possible answers : "Always, Usually, Sometimes, Rarly & Never" items' values are mapped to "1" or "0" in the dataset. For questions 1-9 (A1-A9) in Q-chat-10, if the respose was Sometimes / Rarly / Never "1" is assigned to the question (A1-A9). However, for question 10 (A10), if the respose was Always / Usually / Sometimes then "1" is assigned to that question. If your child scores more than 3 (Q-chat-10- score) then there is a potential ASD traits otherwise no ASD traits are observed.

I did observe statistics such as rate of Class ASD Traits (value equal to 'Yes' / total records), and mean value of the Q-Chat-10 score, which is the sum of questions/values A1 through A10 for the characteristics stated above.

This yielded some interesting findings.

- Males made up 69.7% of the subjects in this study. The percentage of males showing Classic ASD traits was 72.7. The mean Q-Chat-10 score for males was 5.43.
- With respect to age range, the highest average Q-Chat-10 score was 5.64 for subjects aged 27-31 months. The highest rate of Class ASD traits was 77.54% for subjects aged 22-26 months.
- With respect to ethnicity, the highest average Q-Chat-10 score was 7 for subjects designated as Native American and Pacifica. The highest rate of Class ASD traits was 100% for Native Americans. However, only 8 subjects were designated Pacifica, and 3 Native American. If we consider the idea of sample size times probability, and sample size times 1-probability should both be greater than or equal to 10, the following table is useful for looking at sub-categories of ethnicity.

Ethnicity	Value Counts	Proportion Total	Prob ClassASD = Y	ValCount x Prob ClassASD = Y	ValCount x Prob ClassASD = N	Mean Q-Chat-10 Score
White European	334	31.69%	74.85%	250	84	5.72
asian	299	28.37%	70.90%	212	87	4.98
middle eastern	188	17.84%	51.06%	96	92	4.27
south asian	60	5.69%	61.67%	37	23	5.02
black	53	5.03%	73.58%	39	14	5.72

Ethnicity	Value Counts	Proportion Total	Prob ClassASD = Y	ValCount x Prob ClassASD = Y	ValCount x Prob ClassASD = N	Mean Q-Chat-10 Score
Hispanic	40	3.80%	75.00%	30	10	5.60
Others	35	3.32%	82.86%	29	6	5.63
Latino	26	2.47%	76.92%	20	6	5.92
mixed	8	0.76%	62.50%	5	3	4.63
Pacifica	8	0.76%	87.50%	7	1	7.00
Native Indian	3	0.29%	100.00%	3	0	7.00
	1054	100.00%				

At this point, I use SQL to create binary, "dummy" variables for all categorical values. I used the .corr() method in Python to create a table of correlation coefficients. This yielded some interesting results as well.
The correlation coefficient for male/gender to ClassicASDTraits was 0.1177.
The most significant correlation coefficient for age range to ClassicASDTraits was -0.1313 for the 12-16 month age range.
The most significant correlation coefficient for ethnicity to ClassicASDTraits was -0.1815 for Middle Eastern subjects.
The correlation coefficients for Jaundice and the Completed By… categories indicated an even weaker relationship.

Correlation coefficients for behavioral observations A1-A10 indicated strongest relationships to ClassicASDTraits. The strongest of these as follows.

- A5: Does your child pretend?, 0.5633
- A6: Does your child follow where you're looking?, 0.5694
- A7: If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them?, 0.5632
- A9: Does your child use simple gestures, e.g. wave goodbye?, 0.5773

I will examine these relationships further with machine learning and predictive modeling.